

SYSTEM AND METHOD FOR FILTERING A DOCUMENT STREAM

Abstract of the Disclosure

A system for filtering documents and includes a document parser, a profile parser, and a comparator. The document parser accepts incoming documents as input and provides inverted lists of terms contained in the document's output. The profile parser accepts as input user queries and provides 5 as output query nets representing the user queries. The comparator compares the inverted lists representing the documents against the query that is representing the user queries to determine if an incoming document matches a user query. A related method for filtering incoming documents includes the steps of receiving an incoming document and parsing it to produce an inverted list of terms contained in 10 the incoming document. The inverted list is then used to retrieve user queries. Any user queries matching less than a pre-determined number of terms are immediately discarded. The remaining user queries are scored and user queries having a score less than a predetermined threshold are discarded. The remaining user queries are the queries which the incoming document matches.

15

20